# Big Data in the marketing of Finnish B2C companies

Santtu Pursiainen

**Author(s)**
Santtu Pursiainen

**Degree programme**
Business Information Technology

| Report/thesis title | Number of pages and appendix pages |
|---|---|
| Big Data in the marketing of Finnish B2C companies | 44 + 1 |

The goal of this research is to examine the current state of big data used in the marketing of large Finnish B2C(business-to-consumer) companies. The research attempts to analyse whether these companies apply big data technologies in their marketing, and create a generalisation of the current status of the technological advancement on the given sector.

The subject itself is current and is being examined and utilised widely throughout the globe. Different studies are being conducted simultaneously, and the race for the winning solutions is obvious.

The report uses two different studies as a foundation for its base assumption on evaluating the situation and based on them, creates scenarios for the plausible outcome. The scope of the research is limited to non-technological features of big data usage among the participating companies. The report investigates how the companies are using and benefiting from big data in their marketing, and what kind of downsides or disadvantages it possibly brings to the business. Also the level of big data usage has been examined, as well as interviewees´ opinion whether the companies are satisfied with the current level.

In order to validate the assumptions, information gathering for the research was performed by interviewing companies using both quantitative and qualitative methods. The interviews were conducted in person or by phone. The main consideration for the interview method was to conduct it verbally, not as a questionnaire. The interviews were held with people in the organisation with a significant competency in marketing and their utilisation of big data, initially targeted at marketing managers.

The conclusion of the study is that large Finnish B2C organisations are, in fact, using big data in their marketing, some more than others. One of the main benefit of big data usage among the participants was creating better targeted marketing. The outcomes of the research also reveal that using big data in marketing, the companies have a good general knowledge of their loyal and regular customers, but most of them lack the information of their occasional customers, or cannot utilise the information they have. This is mainly due to legislation restrictions and privacy of the consumers.

As a conclusion, the report states that big data is still an evolving technology in Finland which has been in practise for a relatively small period of time. All of the interviewed companies stated that they have room for improvement in utilising the technology, and would do so in the future.

Every respondent variably noted that the future would change the usage of big data and predicted the likely increase in its use. The respondents agreed that new technologies and new ways of using data will be introduced and discovered.

**Keywords**
Big data, marketing, B2C, analytics

# Table of contents

# 1   Introduction

The objective of the report is to introduce the utilisation of big data among the participated Finnish business-to-consumer companies. The report will introduce the main related concepts and create some generalisations of the topic and discuss the key findings. Reading the report will give general information of data, big data and marketing, the description of the other upcoming topics and ideas what are they used for.

The main objective of the report is to examine the current level of the big data usage in the marketing of large Finnish B2C companies, and present the results with a discussion part. The report will dig into the topic and defines on how are the companies using big data in their marketing operations, and whether they are satisfied with the level of their usage. The report will also analyse what are the benefits and downsides of using big data in marketing, and the possible restrictions of the usage.

The theoretical framework will introduce different concepts related to big data and the needed terminology. Topics such as data security and glimpse of the technologies will be looked upon to, and some ideas for the possible results will be discussed. The report uses two previously conducted researches as a base of information, and according to these, creates 3 different plausible scenarios for the results of this research.

The information for the research is collected with interviewing the participated companies, with quantitative and qualitative features. The interviews are conducted in a way that the received information can be considered valid, and furthermore, get a good enough image of the organisation´s big data usage in their marketing.

The empirical part will grasp into the interviews and their outcomes. The questions in the interviews are introduced and analysed, and some suggestions for the results will be made. The empirical part will use graphs and charts to visualise the results among the answers, and create a good base for the final analysis.

The last part of the research will go deeper into the analysis of the whole project, what kind of results can be generalised based on the interviews, and what cannot. The discussion part will put all the previous subjects together and open up the result of the riddled possible scenarios.

The final discussion part will also work as the self-assessment of the research, for the author to explain the project and different phases of it, and lessons learned.

## 2 The research

### 2.1 Research, the objectives and the scope

The main objective of this research is to find out whether big data has been applied in the marketing of Finnish B2C (business-to-consumer) companies. With the main objective, the research studies how have the organisations been able to apply this kind of technology in their operations and especially in their marketing. The study uses two main sources to create a base assumption of the possible results.

The first source is a research conducted by Pursiainen (2015), the topic area covers big data usage in the marketing of Finnish startup companies in the metropolitan area. The result of the research was that startup companies in metropolitan area do not generally apply big data in their operations. The reason for the outcome was widely based on the lack of resources in the use of the interviewed startups. This result suggests that organisations with more resources could be able to apply big data in their operations.

The second source for base assumption was another research done by Alanko & Salo (2013). This research surveyed the present and possible future state of big data usage within different sectors, including consultancy, service providers and schools. The outcome of the research was that the concept of big data is being adopted slowly but will grow rapidly in the future. The research points out that big data as a phenomenon is interesting for every organisation and the potentiality of the technology have been acknowledged.

The two sources used for creating a base assumption for this research both imply in the direction that big data would be used in today´s B2C companies in Finland. This information will be used to create the scenarios introduced at the end of theoretical framework.

The scope of this research is limited to Finland, to large business-to-consumer companies, and how they are applying big data in their marketing. The general big data usage within the companies will be inspected as well, to be able to create generalisation and compare it to their marketing efforts. The theoretical part of the report will create a general view of the global situation and compare Finland´s big data usage in the worldwide utilisation, to create an image of the current state of the nation. The empirical part will only concentrate in creating a view of big data usage in Finland.

## 2.2  Research methods and information collection

The method for the research was chosen to create as valid as possible result with relatively small amount of participants. With the objective of the research, the desired results should create a more profound image of the current situation on big data usage in the chosen companies. The information gathering should be conducted in a way that it leaves room for discussion and further questions if necessary. Due to these requirements of the information gathering, empirical research was chosen to be the most suitable information gathering method. The empirical research is done merely in qualitative manner, but has some quantitative features as well.

The research was conducted with interviewing the participant either in person or via phone. This method allowed the discussion and further questions, which was identified as on the objectives of the research. The interviewees were all chosen to be from large B2C organisations with a fair amount of resources in their use. The pursued persons for interview were preferably marketing managers, or other personnel that had good general knowledge of the marketing situation and their data usage. By choosing this kind of companies and interviewees, the results would be comparable to each other despite the differences in the business sector.

# 3   Theoretical framework

As stated before this research uses information from surveys made by Pursiainen (2015) and Alanko & Salo (2013). This research will be concentrated on business-to-consumer products and services providing organisations that operates in Finland. The case will go deeper in the subject itself; marketing will be inspected and analysed more carefully, Big Data will be introduced more profound as well as the link to Big Data will be created.

The theoretical framework will introduce Data and go on to Big Data, the reader will be familiarized with predictive analytics, as it is one of the main reasons for Big Data, some big data technologies and some theory about decision models. The basics of marketing will also include business-to-consumer sector, and basics of customer behaviour analytics will be introduced.

The last chapter of theoretical framework will associate all the subjects together as they are being handled as one combined research topic. The goal of the chapter is to collect all of the elements of each subject and create some ideas what can be done with the previously introduced subjects. The possibility for such implementation will not be reflected to possibilities in the real business world, but just to create an idea of the underlying potential.

## 3.1   Data

Data has been surrounding us as long as we have been here. Data is simply everything we can comprehend; data is written or observed information without deeper meaning at this state. Data can be numbers, letters, sound or whatever our senses can collect for our brain to process. Only after we can understand the received data and analyse the meaning of it, it becomes information. Data can be collected in multiple ways for many different platforms. But the main idea is that it is being collected. As Arthur (2013, 48) points out, every time we use technologies in today´s digitalized world, we are creating data, which can be collected, stored and used. But only collected and stored data can be reliably used afterwards, since there is proof of it being existent.

Data itself does not bring too much to the table, after understanding the meaning of it, it becomes information and so on. Wu (2000) suggests that data is the first step of the 5 in transforming data into wisdom. Data has been around a long time and the description provided in 2000 can be thought to be valid today as well. The steps for the transformation are:

1. Data
2. Information
3. Analytics
4. Knowledge
5. Wisdom

In these steps Data is the raw material, which have been possibly collected from different sources. It can be numbers, text, voice or for example click streams. This kind of data can turn into information when it is possible to make some kind of relation between the pieces of data. Information is data which can be interpreted somehow. Analysing the data could be connecting two pieces of data, a simple example, a series of numbers which can be identified to be a phone number, and a name of a person and other information such as occupation and salary. Combining the information will create contact information. When the information has been analysed, it can be stored somewhere and accessed by others if necessary. After the stored information has been understood by a person, it becomes wisdom. In this simple case the wisdom would be knowing this phone number and using it to contact this person and maybe use it for sales opportunity. (Wu 2000.)

The goals of organisations might not be as simple as the example, they might pursuit different kind of knowledge and wisdom regarding their business operations or customer segments. This kind of transformation path might able creating personal views of customers and their preferences as a consumer, or optimizing a process to increase efficiency.

## 3.2   Big Data

Salo (2013, 28) suggests that big data as a concept refers to data, but also to the services, products and technologies that enables organisations to deal with the challenges of the data. As Franks defines (2012, 7), big data differs from traditional data in the way that the data is automatically generated by machines, not by humans. This allows the high amount of data to be collected in such a rapid fashion. From which we come to the definition of big data: (Arthur 2013, 47; Finlay 2014, 13; & Salo 2013, 21) all agree that Big Data is data consists of three V´s, which are:

- Volume
  - Volume of the data requires large databases to handle large amounts of already existing data and capacity for new received data. Finlay (2014, 13) suggests that data becomes Big Data after it cannot be managed on normal PC or laptop.
- Variety

- o Variety means different types of data. This kind of data can be structured or unstructured, it might be in the form of text, voice, image or collected clicks in a web advertisement for example.
- Velocity
  - o Velocity means the high speed of incoming data that might change over time in some cases, for example someone´s address or marital status. This kind of data is being generated continuously and has to be accessed within all the data and include the possibility to be changed. Usually this kind of data is fragile and has to be well handled not to become corrupted or even lost.

As many sources define big data consisting of the 3 V´s, although this is just one part of the definition, Gartner (2016) suggests that it is also about "information assets that demand cost-effective, innovative forms of information processing that enable enhanced insight, decision making, and process automation".

Now it is known what is big data, but what does it do? Why should it be collect large amounts of data in all possible forms and just store it somewhere? It should not. The main idea of Big Data is to use it, and understand how can it be used, otherwise the whole idea goes to waste and there are just huge loads of data in one´s hands. Big Data is only valuable when the information is being harnessed for analytics, decision making, or something else an organization can benefit from. All of the collected big data will not be valuable for the business, but the valuable part has to be found and then used. (Franks 2012, 17.) According to Salo (2013, 28), the three main challenges of big data are the storing, transmitting and analysing of the data; how to gain an understanding of the past, present and future state of events based on data.

As Finlay (2014, 18) believes that the most essential usage of big data is to create new forms of customer predictions feasible. This means creating new channels where to collect data or different way to analyse the existing one. It is for creating new and better ways to predict the upcoming events. One form of big data according to Gartner (2014) is dark data, which is big data that has been collected, processed and stored, but has not been utilized to benefit the business. Dark data could be pieces of information among others, that have not been recognized beneficial or failed to been noticed altogether.

Press (2013) introduces multiple different publications that are talking about the increasing amount of data, and the term "big data" has been used even before the 21th century, most of these publications are originated from the United States. The usage of big data has concentrated in USA more than anywhere else, which makes them the forerunners in that field of technology. According to Vaakkuri (2013) Europe is 2-3 years behind U.S. in the usage of Big Data. Which is a lot in today´s world. Gartner´s research, Gartner (2015), addresses that the "hype train" of big data has gone, but still within two years 75 percent of

the companies are investing or planning to invest in the technology, as it will become the new normal. Salo (2016) agrees that the peak of the hype has already has been reached, or will be reached during this year, and the concentration is on the actual utilization of the technology.

### 3.2.1 Data storage

The wanted data has to be stored somewhere to be able to use it. Today there are 2 main solutions for data storage; Physical, more traditional way to store data, or cloud computing solutions. The actual storing of data is accomplished in a physical database in both of the solutions. The first option means that the organisation has their own database located somewhere in their facilities, and cloud computing is using a third party to store the data. Both of the solutions can include similar data, in other words, all of the data that the organisation has. As Hovi, Hervonen & Koistinen (2009, 4) declares, the most of the needed data is stored into organisations´ own databases. This might have been true in 2009, but since then cloud computing has vastly grown. Salo (2010, 16) suggests that cloud computing refers to a service provided by $3^{rd}$ party, which offers the usage of software, equipment or services via internet. Which means to be able to use these functions and tools without having to worry about the maintenance or other up keeping responsibilities. In other words, the data is stored in someone else´s database and the company can access and use the data via internet.

### 3.2.2 Predictive models

Predictive models are different ways to analyse data to predict something that will happen in the future. They are based on existing data, and relying in cyclic occurrence. Predictive models are based on using different factors on events to create image of the present situation and based on that, forecast the upcoming events.

Finlay (2015, 105) suggests that two different types of data are needed to create a predictive model; predictor data and behavioural data. Predictor data is the collected, known data, which will the basis of the model. Behavioural data is the outcome of the prediction after applying the used variables to the predictor data, the forecast. Behavioural data tells the behaviour of the customer, which is one of the goals of big data.

There are as many predictive models as there are users, the models can be used differently and anyone can create one. Although not all of them are equally good, or works on every occasion. Finlay (2015,104) implies that the most commonly used predictive models are:

- Linear models
- Decision trees
- Neural networks
- Support vector machines
- Cluster models
- Expert systems

The two most popular ones are linear models and decision trees, which will be introduced next as an example.

**Linear model**

As stated before, different models have different kind of features and variables to work with. Finlay (2013, 106) suggests, that in linear models relationships between predictor variable and behavioural are produced by weight. The final score is the outcome of the predictor variable multiplied by the weight and added to the behavioural score. Then the scores are placed on a chart, which will create the mean value, line between the digits used.

**Number of ice creams sold**
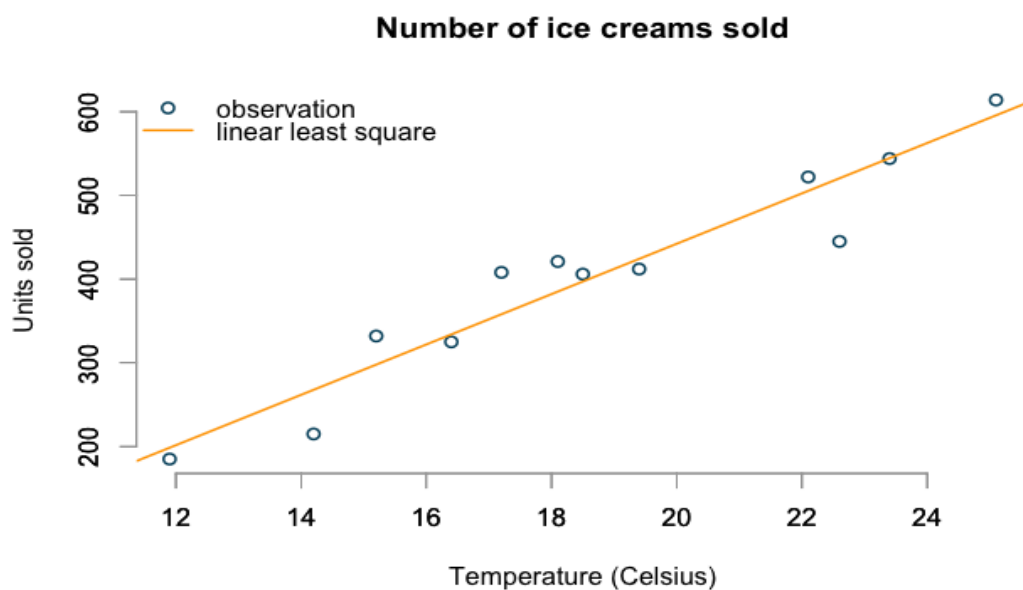


Figure 1.1 Linear model of ice cream sales compared to temperature (Gessman 2015).

In this example, the measured object is ice cream sales compared to temperature. The observations are the circles and the line represents the linear mean of the observations. The chart tells more ice cream is sold when the temperature goes up, which would be expected, therefore we can predict every time the temperature increases, ice cream sells better. That is a simple prediction.

**Decision tree**

According to Finlay (2015, 112) decision trees are probably the second most popular type of predictive model used today. The name comes from the possible representation as in the figure of a tree. There are two different tree types, classification and regression trees Finlay (2015, 112).  Classification trees are mostly if-then based, and often referred as the default decision tree. Regression tree is used when the response variable is numeric or continuous, like predicting consumer product price. The decision trees are based on segmentation, quite often population and their behaviour. Decision tree creates a segment and then breaks the segment into smaller ones, the segments can be created based on income, number of members in a family, owned cars by family or whichever is the measured object. Decision trees can also be true-false based, and lead into a decision after answering a number of questions. The outcome of a decision tree could be whether to invest on something (does it probably create profit) or how many cars are there averagely in a family of 4 members. Therefore, decision trees can be used in other than making direct decisions too, but is merely used to gain information base in decision making. (Deshpande 2011.)
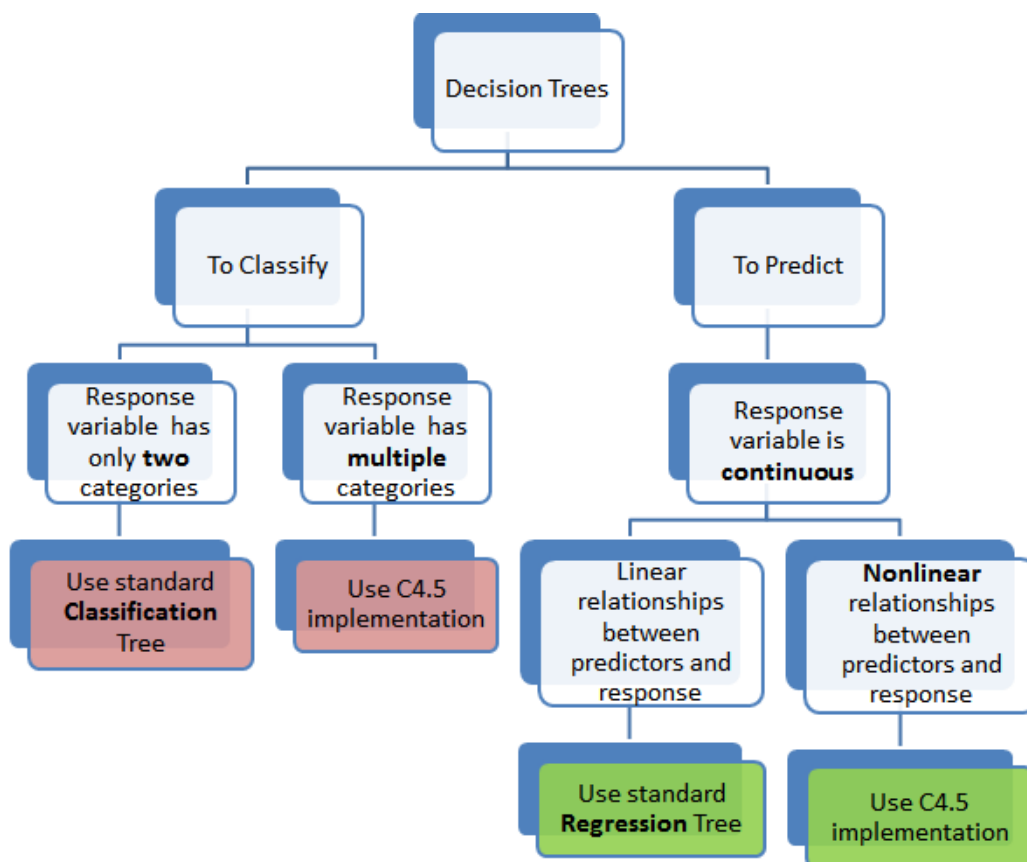


Figure 1.2 A capture from Deshpande´s blog, which introduced the two (classification and regression type trees), and in which situations they should be used. C4.5 implementation

refs to an algorithm used to create decision trees with more variables, which will not be introduced in this research. (Deshpande 2011.)

### 3.2.3 Predictive analytics

Predictive analytics is one the main ideas and usage of Big Data. Predictive analytics come from predictive models used to create continuum of events. Events tend to repeat themselves in almost any kind of occasion. By analysing data, there is a good chance the future events can be predicted based on the past. A simple example would be grocery sales; a grocery store collects data that mincemeat is sold the most on Mondays, so the highest point of sales for mincemeat is probably on future Mondays too. Although this kind of events, usually a bit more complex, can be predicted by using data. Another slightly more complex example could be: what is the customer most likely to purchase with the mince-meat? Or what is the age of the customer, or does he/she have a family? This kind of questions might need big data to be able to tell the answer. The bigger the industry, the more variables there are, and the amount of data is being created simultaneously.

There are numerous ways to use analytics and base business operations on data, according to research done by (Press 2015) 51 per cent of the executives feel that adapting and refining data-driven strategy is the single cultural barrier. 47 per cent of the same executives felt that putting data learning into action is a challenge. Therefore, the research suggests that decisions are not based on data widely enough.

If considering about customer predictability, it is mostly based on behavioural models. How do the customers act, when do they visit stores and when do they purchase something? Finlay (2014, 68) suggests that 7 different data types are needed to predict one´s behaviour:

| Data type | Description |
|---|---|
| **Primary behaviours** | Information about a past behaviour that is of the same type as the behaviour you want to predict. If a person has committed burglary before, they are likely to do it again. |

| | |
|---|---|
| **Secondary behaviours** | Information about past behaviours which are similar but different to the predicted behaviour. Littering and non-payment of parking fines are not major crimes, but they indicate a propensity to commit worse offences. |
| **Tertiary behaviours** | Information about past behaviour that does not have an obvious or direct connection with the behaviour being predicted. If someone has sold a lot of items on eBay recently may give some bearing on their propensity to burgle. |
| **Geo-demographic** | Details about a person´s state of being, rather than what they have done in the past. For example their age, blood pressure, weight, hobbies, what car do they drive and occupation. |
| **Associated data** | Information about other people, someone has a connection too. This includes the associate´s behaviours (primary and secondary and tertiary) and their geo-demographics. If someone´s spouse has committed a burglary, there is a greater propensity for them to commit burglary as well. |
| **Sentiments (feeling and opinions)** | A person´s attitude to things: their likes/dislikes, approvals/disapprovals. If someone answers a crime survey by saying they think it is OK to steal, or they write a derogatory blog about law enforcement, then they may be more likely to engage in criminal activity than the average individual. |
| **Network** | This is information about the nature of the connections between a person and their associates. If someone has 100 friends and |

| | 99 per cent of them are burglars, then they are likely to be a burglar too. |
| --- | --- |

Figure 1.2 Finlay (68, 2014) illustrates the data types needed to predict the possible behaviour of an individual. The figure uses a burglar as an example and gives examples of where the data could be collected from and then interpreted.

### 3.2.4 Security and legislations related to big data

As it is known, information is stored somewhere physically. In today´s world it is done electronically, and could be always accessible from anywhere. Which means the data can also be accessed by people who should not be able to do so, by hacking into the systems and databases of the organisations. This has created a major problem with security. The organisations try to secure their data as well as possible, but unfortunately the ICT business is rapidly changing field, also among hackers. Since data is so valuable, the data breaches create huge losses for the organisations who loses it. It might cost them a lot in sense of money, and even more severe loss in the long run, the trust of customers.

Salo (2014, 51) examines Big data security from 3 different angles; Information source security, physical security and juridical security.

1. Security for source information
    - Usually the raw data is not much worth when stolen, it might not tell anything and therefore cannot be used to harm a business or benefit third parties. The value of the data comes after processing and analysing the data, the analytics and insights of the data are the ones that are vulnerable to be pursued by hackers. Also this part is vulnerable for the customer as well (Salo 2014, 51-52.)
2. Physical security
    - Databases contain a lot of data and require high speed processing and might be processed with such (open-source) tools that prefer data which is meant for everyone. Such data is usually stored in a non-encrypted form. This will create a problem to start with, if the data would be accessed somehow by any unwanted party. In this case, the data would be available and in easily readable form. If the data would be encrypted, it would not be readable if somehow gotten into wrong hands (Salo 2014, 52-53.)
3. Juridical security
    - The rights for collecting, storing and using of data is a discussion of today. Who can collect and use the data of individuals? And what kind of data can the organisations or governments collect? Today the legislations are going into the direction of individual decision making power. Where individuals can choose what data is stored from them, and decide if the data needs to be deleted. Therefore, individuals could be anonymous in certain extent. On the other hand, a lot of services for consumer identification are being created, which tips the scale to completely other side (Salo 2014, 54-56.)

Today legislation is lacking behind with new possibilities of big data. Salo (2015, 36) suggests that in the present state it is not certain which kind of information is legal to be collected and which is not. This might result better or worse situation for the future of big data, if the legislation gets stricter, limitations for collecting big data might complicate the situation. On the other hand, since organisations also need the permission from the customers to collect specific kind of data, problems might occur if the permission is not granted. In such cases organisations would not be allowed to collect and use this particular customer´s information.

### 3.2.5  Big data in Finland

Salo (2013, 88) suggests that in Finland big data is thought to be only for the large organisations who has huge masses of data in their possession, and are able to put a lot of resources into the data processing. As Salo (2013, 90) predicts that an organisation which has a good business plan that includes customer need analysis, will succeed in the future. Although the author implies that the innovations of the world are too rarely heard in Finland. The research done by Alanko & Salo (2013) suggests that big data is being used in Finland and the usage is rapidly growing.

Salo (2013, 92-94). recognises 3 factors why Finnish companies are not applying Big Data effectively enough in their business.

1. The urge
    a. There is no significant urge to get to know and try new things that come from the outside world, usually the responsible ones are the top executives of the organisation that does not understand the need of constant change.

2. Technology
    a. Technology is not a problem generally. Different solutions are easily available in a low price. The usage just has not met the availability and the possibilities.
3. Know-how
    a. The knowledge is a problem; new technologies are not taught in schools or they are not implemented in organisations. Not enough investments are made towards using new technologies and thus being in a more competitive position in global industry.

As mentioned before, Big Data usage is best practised in the U.S. and Europe come couple of years behind. In Europe in the other hand, according to Slocum (2013) U.K. is the winner of using big data, with the most utilisation of the technology.

## 3.3 Big data technologies

This research does not consider big data technologies itself too closely, and excludes them completely from the empirical part. Some solutions will be discussed in light manner, comparison will be made between cloud and physical solutions, but no detailed information about technologies will be included. Some differences between vendors and solutions are presented in the next chapters, to create an image what kind of tools there are available in the industry and what makes them different.

### 3.3.1 Hadoop based technologies

As both Finlay (2014, 197) and Salo (2013, 80) suggests that Hadoop is one of the most pivotal piece of technology today in the field of big data. Apache Hadoop is an open-source software developed as mass storage system which is used to store and analyse large amounts of data. Added to the low cost and easily accessible features of Hadoop, it also provides reliable data storage.

The two main pieces of software inside Hadoop, the ones Hadoop does not do much without with are Hadoop Database Core tools (The Hadoop common) and HDFS (Hadoop Distributed Filing System). Hadoop common contains libraries and utilities needed by other Hadoop modules. HDFS creates multiple stored piece of data on multiple servers, which requires more storage capacity, but guarantees the safety of the data (Finlay 2014, 201.)

Although Hadoop is very powerful tool dealing with Big Data, there are some limitations or preferences in the processing of data. Finlay (2014, 202) suggests Hadoop works best in circumstances where the data has to be written once and read multiple times. Which means Hadoop can process a large amount of data one time, but not small amounts of data many times. Hadoop´s best usage is to analyse old data to create an image of the existing information, but not to do real time processing or analytics.

Since Hadoop is an open-source technology, a lot of different products are made by different organisations, to meet their own needs and preferences. As Salo (2015) reports in his Big Data blog that the most well-known providers and their solutions for Hadoop based distribution technologies in alphabetical order are:

- Amazon (EMR)
- Cloudera (CDH 5.3)
- Hortonworks (HDP 2.2)
- Microsoft (HDInsight)

- IBM (IHC)
- Intel (Intel distrivution for Apache Hadoop, *exterminated*)
- MapR (M3, M5, M7)
- Pivotal (Pivotal HD)
- Teradata (TDH)

Hadoop is vastly used and well-known technology, but the big players invest a lot on their big data solutions and do not use open-source software (although some components might be based on them). The ones who have costly and vast solutions, are for example IBM, Oracle, Microsoft, Hortonworks et cetera. Vendors like Oracle and IBM are more relevant for national and international larger organisations. A lot the international vendors use different solutions for different operations, with such amounts of data the operations has to be distributed with multiple sources. Databases have their own solutions, and data management is handled with a different system. The vendors might offer Hadoop based solutions and also their own products which are created differently. As it was said earlier, Hadoop is good for big data sets for write once and read many times.

### 3.3.2   Apache Spark

Apache Spark is another open-source framework for handling big data. It has similar components as Hadoop does, but the functionality is different. Spark differs from Hadoop in its processing abilities and functionality with big data. Spark´s advantage over Hadoop is the speed. Spark has been built to process data and create analysis in near real-time. Hadoop uses large data sets at once, and enables the reading of large amount of data, which is stored in the database. Spark, on the other hand, brings smaller data sets to the faster logical RAM memory, and enables fast processing of data as they are in smaller sets and located "in memory". Spark is good for real-time analytics such as recommendations in a web store or in a search engine. (Marr 2015.)

### 3.4   Marketing

The main idea of marketing is to provide information about your products to the potential customers. Marketing can be done directly to a specific person or to masses trying to reach as many as possible at once, and everything in between. As Fahy & Jobber (2012, 5) define, marketing is about meeting and exceeding customer needs better than the competitors do, and make it as a corporate goal. Three main points have to be achieved before applying marketing concept according to Fahy & Jobber (2012, 5.):

1. Customer orientation
   a. The organisation should be concentrating on creating and increasing customer satisfaction. This means that the organisation should be able to

concentrate on their customers as well as they do on the staff or the products produced. By doing this, the contact with the customer can be well preserved and meet the changing needs. If this is not in the priority list, customers may be lost due to the lack of satisfaction from their behalf.

2. Integrated effort
    a. The whole organisation should be part of the marketing efforts and preserving the customer satisfaction, not just the marketing department. If the whole company will put effort on giving a good and caring image towards the customers, they are more likely to succeed in creating the brand they are aiming to.

3. Goal achievement
    a. Achieving goals are highly dependent on psychological success. The organisation has to believe that the set goals can be met, and create an atmosphere that everyone else does too. Achieving goals is impossible if the people working for it does not even believe it will happen.

As stated, marketing is about meeting and exceeding the customer satisfaction. Although the information has to be collected first, the information does not create itself. The organisation has to use all possible channels to be able to create an image of their customers, what kind of products do they purchase, when or where do they purchase them and so on. When the organisation has knowledge of the customer, they might be able to figure out what are the preferences of a specific consumer. According to Tollin & Caru (2008, 58), market impact can be identified with different measures such as the number of times a website has been visited or a product has been purchased, and to create an estimate of for example the product diffusion and the demographic distribution.

### 3.4.1 Customer value

As stated in the previous chapter, marketing is about customer satisfactory. Which means the value of the customer is undoubted, critical to the organisation. According to Fahy & Jobber (2012, 6) customer value comes from perceived benefits is larger than perceived sacrifice. So no more should be sacrificed to the customer, than the customer adds to the business of the organisation. Weinmann (2015) states that customer relationship is more valuable than independent transactions, because the customers tend to repeat their actions. By doing so, customers are an endless source of income to the organisation. Customers bring value to the business according to their level of satisfaction.

The value of the customer can be increased in multiple ways, such as campaigns, advertisements, competitions or something else that would catch the customers' attention. The whole marketing efforts are balancing between the previously noted equation, when does the effort create more benefit than sacrifice resources.

### 3.4.2 B2C

Marketing process can be aimed to different fields of businesses. In this research the aim is on consumer products and services. Business-to-consumer products means products and services that an organisation provides to end-user consumers, rather than other businesses (Marketing terms 2016). Considering big data and marketing, the consumer products offer more relevant data and more variables due to the large amount of customers, than businesses would. Which is the most centric reason for the chosen scope, also regarding the readers. The consumer products are easier to grasp and relate to as a consumer, than any business-to-business products would be.

### 3.4.3 Consumer behaviour

As Schiffman & Kanuk (2010, 23) defines consumer behaviour as the "behaviour that consumers display in searching for, purchasing, using, evaluating, and disposing of products and services that they expect will satisfy their needs." Today consumers are more aware of the assortment of products and perhaps more demanding than ever. There are so much information and different products to choose from, and the products are easily purchasable. Consumers can easily compare the prices or models over the internet and order them straight to their doorsteps. This is why the companies of today have to be able to compete with every single one in the same field of business, and offer something that the others do not.

This subject will be tackled by dividing marketing into 3 segments: Market segmentation, targeting and positioning Schiffman & Kanuk (2010, 28). Market segmentation is about identifying the targeted consumer groups whom the organisation will market their products to, and also build up they brand for. As Schiffman & Kanuk (2010, 76) defines, the segmentation can be consumer-rooted or consumption-specific, and the information will be based on facts or cognitions.

Consumer-rooted facts are pieces of actual collected information about demographics such as age or gender, or purchase history, numbers that can be measured. This information can be used to categorize the consumer based on the used attribute. On the other hand, consumer-rooted cognitions are information about behavioural measurements, which cannot be accurately diagnosed. There is no way so far to exactly measure and predict an impulsive buyer for instance, or know personality traits of a consumer, such as confidence. (Schiffman & Kanuk 2010, 76.)

Consumption-specific information includes the same two sides of information than consumer-rooted information did, which are facts and cognitions. In consumption-specific behaviour, usage behaviour examination includes usage rate and usage situation Usage rate segmentation divides consumers into heavy, medium, light and non-users of a product or service (Schiffman & Kanuk 2010, 88). These usage rate segmentations are purely based on the amount of purchases made by a consumer. As Reh (2016) explains, in many cases 20% or the consumers create 80% of the sales. Which means a lot of the organisations try to identify the heavy users and market their product to that segment, who create the most of the sales.

Other segmentation is done based on usage-situation. Which is highly based on holidays and such. Usage-segmentation is about trying to predict the upcoming high selling points of specific products, and in this case, it is events. A consumer will most probably purchase something for their family members´ birthdays or anniversaries, maybe on valentine´s day and on Christmas. If the organisation knows about the consumer´s children´s birthdays or the couple´s anniversary, they can directly market products to fit this occasion. (Schiffman & Kanuk 2010, 89)

The consumption-specific cognitive segmentation is divided in benefit segmentation and brand loyalty and relationship (Schiffman & Kanuk 2010, 88). Benefit segmentation is about what can be added to the consumer from the occurring purchase, or what does the consumer need. For example, if a consumer buys designer jeans, the person might get some social benefit from it by creating a certain image of themselves. On the other hand, a consumer might get more value for money if they buy good quality non-designer jeans with much less money. Brand loyalty and relationship is about the consumer´s behaviour and attitude towards a brand. In this segment the organisation is trying to analyse whether loyal and regular customers bring more to the business than occasional shoppers do. (Schiffman & Kanuk 2010, 92-94.)

The segmentation can be done in different ways, for example using algorithms for creating patterns and consumer behaviour. Grigsby (2015, 65) introduces a calculation that uses education as a factor for response rate for distributed catalogues. The simplified result for the formula was that; the more educated a person was, the more probable he/she was to purchase something from the catalogue. This kind of results would suggest to target high educated households to get the best possible response rate for the catalogues.

## 3.5 Combining the subjects together

### 3.5.1 Data driven marketing

This is the part where Big Data and marketing starts working together. As stated earlier, Big Data is large amounts of data chunks which are being collected, stored and analysed. Marketing on the other hand is about customer satisfaction and knowing the customer. It becomes quite obvious why these two should be used together. Arthur (2013, 48) declares that data driven marketing is about collecting, analysing and executing big data to increase customer engagement, enhance marketing development and measure internal liabilities.

What could be a better way to create a perfect 360-degree view on a customer than a huge amount of information on their previous purchases, locations, times of purchasing et cetera. Of course the information has to be collected first, but let us assume that the organisation has done just that. When the organisation has explicit information of the consumer preferences, it should be easy to manage the product quantities proactive to time and place. By doing this grievance could be avoided and customer satisfaction held up.

### 3.5.2 Customer insights

As Finlay (2014, 18) implies that the finest opportunity of big data is to enable customer predictability. By being able to predict customer behaviour, the organisation can for example perfectly recommend the products the customer needs, even before they ask about the product. The customer might not even know that they need something before the product is recommended to them. In cases of recommending something similar to the purchased or viewed product, the customer might realise that the recommended product is essential to the existing one, and purchase it too. A very valuable prediction could be made on health, if the upcoming deceases or conditions could be predicted and healthcare could be offered before the danger, lives could be saved. So the customer received a recommendation of a product or a service that the person was unaware of at first, or did not have the urge or need for it beforehand. (Finlay 2014, 18.)

### 3.5.3 Target marketing

As Fahy & Jobber (2012, 122) suggests, marketing segments have to be identified before target marketing can be applied. The criteria according to Fahy & Jobber (2012, 122) for the defined segments are:

- Effective
  - The effectiveness of the segment can be measured from the needs of the customers; each segment should consist of rather similar interests of the

customers. Although the interests of the customers in this segments, should differ significantly from another segment, to be able to differentiate the segments clearly.

- Measurable
  - o The customers in the segment needs to be clearly measurable. This requires distinct characteristics of behavioural patterns. This kind of factors could be age, occupation or marital status.
- Accessible
  - o This criterion could be identified as knowledge about which kind of promotional campaign of products are effective on which groups of people, and who the product might concern or interest.
- Actionable
  - o Actionable can be simplified as the budget to execute an opportunity of target marketing. There is no point on planning to reach onto a set of potential customers if the initial condition is that there will be no money to do so.
- Profitable
  - o As expressed as the most important criterion of segmentation, will it be profitable? The segmentation needs to bring more money to the company than the segmentation will consume. There might not be any point of segmenting the least income creating group and marketing products to them.

After the segmentation has been done, the marketing can be aimed to those specified groups. Now the organisation knows what kind of segments they have, what are the customers` preferences in that specific segment and their potentiality of bringing profit to the organisation.

Fahy & Jobber (2012, 123-126) implies that there are four main strategies in target marketing:

1. Undifferentiated marketing
   a. This type of "target" marketing is when you do not have a target. This can be considered as the default level of marketing. In this strategy the customers are not well known enough to segment them, or the segmentation effort has not been instituted.
2. Differentiated marketing
   a. Differentiated marketing is the opposite of undifferentiated. The market segments are defined and they have been taken into consideration in marketing efforts. In this strategy, all of the defined marketing segments are implemented and different target marketing campaigns and product lines are created.
3. Focused marketing
   a. The type of marketing that has defined all of the segments of marketing of its products just like in differentiated strategy. Although, in focused marketing, the company only uses perhaps one of the segments to use as their target. This kind of marketing is usually practised by small companies with limited resources, or by a company that has something very specific to offer.
4. Customized marketing
   a. Customized marketing is about having a basic product which can be customized for the customer. These kind of product´s can be mobile subscriptions, architectural services or sneakers. In these cases, most of the of customization happens on a customer-by-customer basis, which

traditionally requires a presence of a person to receive the requirements of the customization. Today, more and more of these customizations can be done via internet, and the need of interaction with another human being is becoming irrelevant. This enables mass customization marketing, which will allow customers for example design their own product in the stores´ websites. The future of marketing is to increase the amount of mass marketing in a customized manner.

## 3.6 Research question and possible scenarios

The research is covering area around big data, marketing and analytics. The research brings these all together and creates a view of what is happening on this field of technology today and what might it lead to. Although the main research question lies within big data, does it get enough attention in marketing, or why should it? Pursiainen (2015) found out that startup companies in Helsinki region do not apply big data very well in their business. The study suggested that big data was not applied in startups due to lack of resources, legislations and lack of data. This research is concentrated to larger organisations with more resources and large amounts of data in their possession. Also the research conducted by Alanko & Salo (2013) suggests that big data is being applied generally. Therefore, the base assumption of the research is that larger organisations use big data at least on some level, but examines whether it is used in marketing. Also the research is trying to find out what is the atmosphere around big data, how is it viewed from the point of marketing departments in large B2C companies, and how well has the idea been adopted.

After all, there is the possibility that a company does not apply big data in their marketing, or other processes, so one possible outcome of the research could be that there are no operations related to big data. Another, more plausible outcome would be that the organisation uses big data but not in their marketing. The third option would be that big data is adopted in the marketing of the organisations, perhaps together with other processes. There could be one more scenario, which would divide the usage of big data in marketing and in other processes, but the distribution of the results might be too wide due to the size of the sampling. Although the most weight has been put on researching the big data usage in the marketing efforts of the organisations. Therefore, three scenarios could be optimal for this particular research:

1. Scenario
   - Big data has not been generally adopted by large B2C companies in Finland
     - Why does not the company use big data in their marketing?
       - the reason; resources, lack of data, lack of expertise, legislation or something else?
       - what are the benefits of not using big data?

- what are the downsides of not using big data?
2. Scenario
    - Big data has been applied but not in the marketing of the companies
        - what is the reason it has been applied to other processes but not in marketing?
        - Will it be applied in the future?
3. Scenario
    - Big data has been applied to marketing and perhaps on other processes too
        - What does the big data bring to the company, benefits and downsides?
        - What is big data used for?
        - What is the level of big data usage today and should the company be satisfied with the current level? The future?
        - What does big data require from the company?

# 4   Carrying out the interviews and presenting the results

The empirical part of the research will be done based on the theoretical one. The empirical information gathering will be executed by interviews; the interview questions are based on the discussed topics of the theoretical framework. The interviews tend to be flexible to enable more open discussion with the interviewees, to create a more truthful image of the situation, rather than easily interpreted yes/no answers. As it was stated in the theoretical framework, the survey is done both in qualitative and quantitative manner, with closed (yes/no), multi-choice questions, and open questions that will create wider answers. The reason for the mixture of methods is to create as wide understanding of the situation as possible with an interview averagely less than one hour in duration. Although this kind of interview technique will leave some room for interpretation mistakes done by interviewer and have an effect on the end results.

The main target of the interviews´ are to create a general view of the state of big data usage in the involved organisations and generalise the results to cover most of the Finnish industry, if possible. If the results vary a lot from each other, a possibility of not being able to create any generalisation based on these interviews might become relevant. The research tries to bring out how big data is used at the moment and is it enough in today´s world. Furthermore, does the companies see any benefits of using big data in their marketing, or do they gain benefit from it.

The interviews will hopefully bring out whether big data has been accepted as a concept in the organisations and their view on the importance of it. Also it should demonstrate the level of big data usage in the organisations, and whether the companies see´s it comprehensive enough throughout the whole organisation. Since the interviews include open questions, the interview methods were chosen to maximize the outcome. Interviews will allow deeper discussion on the topic and include follow-up questions, to gain a realistic image of the actual situation. The questions were sent to the interviewees in advance, but could not be answered as a questionnaire. Although the interview could have been done in person at sight or as a telephone interview.

The problems with carrying out the interviews were related to the size of the organisation and the position of the interviewees. Most of the chosen companies were able to be contacted, and only a small amount declined immediately. However, the problem was in the interview arranging and scheduling, most of the companies asked for further information via email, but it took a while for them to reply and start the negotiating about the time of the interview. I targeted marketing managers of each company, but as the subject of the

research was highly weighed on big data, sometimes the interview was done with someone who is more familiar with that particular subject. This was one of the factors that increased the time of being able to arrange an interview with the company.

## 4.1 The participants

The participants were large and well-known organisations in Finland, and the interviews were conducted with people who are familiar of the general situation of marketing in the organisation and have knowledge on their big data technologies. The participated companies are quite well known for Finnish people in general, but since the research is done as a thesis for international study program, some additional information of the company´s field of business will be provided. From all the 14 companies contacted, interviews were conducted with people in the following positions in the companies of:



Airline

Business analyst in commercial business unit



Fashion goods providing chain

Marketing manager



Grocery store chain

Sales director



Financial service group

Senior segment marketing manager (Nordic)



Railway company

Head of customer insight

## 4.2 Research results

The research results discover the most suitable scenario for the research based on the findings made in the interviews. The interview information will be introduced as the number of the participants and response rate. The actual interview answers will be presented visualising the results by graphs and added clarification.

### 4.2.1 Scenario selection

The research found out some differences between the interviewed companies, although there were a lot of similarities too. Perhaps the different industries of the companies might have done some of the differences, in some cases. The results for big data usage in marketing took majority of the total answers, but yet it was only 3/5. The other two, who answered that they are not applying big data in their marketing, did use a large amount of data in their marketing which they just did not characterize of being big data. Then again, 4 out of 5 of the companies are applying big data in their operations in general. Only one of the organisations were not using big data at all, although they are applying the technology in near future.

With this information gathered, the scenario can be already selected. Since the last 2 scenarios suggest, big data is used but not in marketing (2) and big data is used in marketing and possibly in other processes too (3). The results strongly suggest that the 3rd option should be chosen, which is, big data has been applied to marketing and perhaps on other processes too.
.

### 4.2.2 Interview information

The final amount of interviewees was not as high as planned, the total amount of contacted companies was 14. From the total 14 selected companies:
- 11 were reached by phone
- 9 were interested, and further information was sent
- 2 declined to take part
- 5 actual interviews

Altogether more than 1/3 of the original 14 companies were interviewed, which is rather good rate considering the size of the companies. The first desired amount of interviews was set to 10, which was quite optimistic regarding the amount of contacts. The optimistic starting point appeared to be quite realistic after all, since the amount of the interviews could have been higher with the same amount of contacts if the schedule would have been different. Although the amount was short from the original desired 10, the research can be considered to be reliable.

### 4.2.3 Results of the questions

**Question 1. What do you consider being big data?**

This question was to create a starting point for the interview and to make sure we have mutual understanding of the subject. When given the opportunity in each interview to

explain what they considered to be big data, the results differed from each other slightly. One characterised big data to be a problem statement rather than just data or technology, another interviewee suggested that it is more philosophical view what can be considered being big data. Although everyone agreed that it can be characterized as structured and unstructured data coming simultaneously from different sources in high velocity. This created a mutual understanding of the discussed topic and did not vary in any of the interviews:

- Structured and unstructured data that is being generated continuously from different sources in large masses
- Information about customers and the ability to this data to create a better view of the customer and discovering their needs.
- Data processing and handling, and utilizing data in solving problems.

**Question 2. How would you describe the usage level of big data in your company?**

     a) Very good
     b) Good
     c) Rather good
     d) Rather basic
     e) Bad
     f) We are not using big data
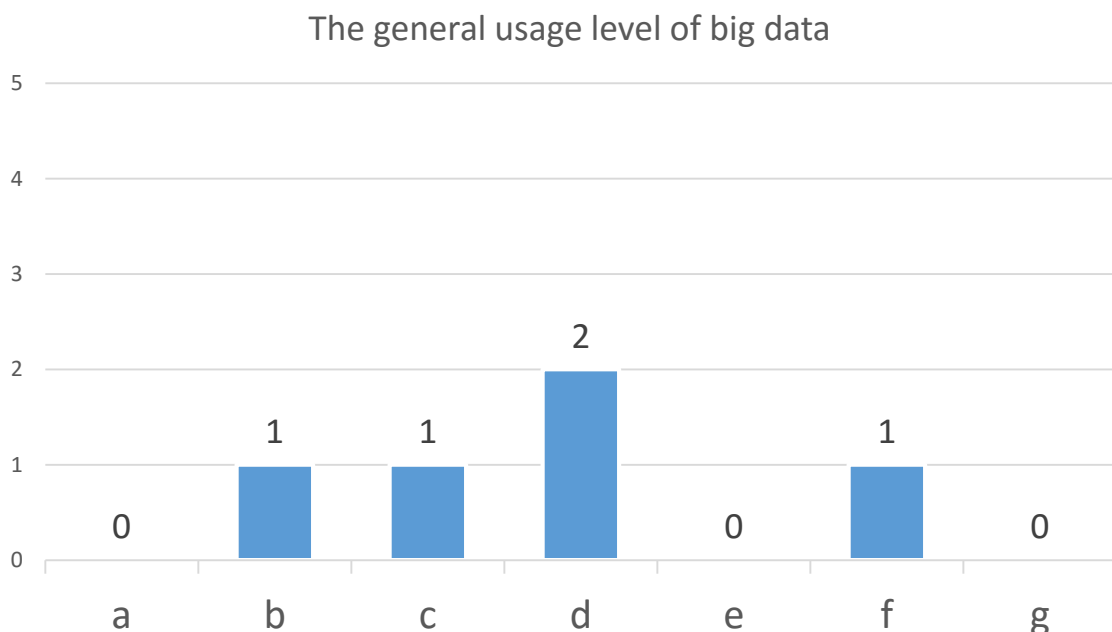     g) something else, what?



Figure 3.1 The graph shows the distribution among the usage level of big data in the interviewed companies. The axis on the left side indicates the number of participants and the axis in the bottom indicates the option described above.

As seen in the graph, the slight majority (2/5) of the participants choose d, one participant chose c, one b and one f. The distribution between these answers are quite vast, but in most of the companies´ big data is used in at least decent level and the usage is growing

towards the future. One exception was the one organisation, which did not apply big data but expressed their data usage being on rather basic level.

**Question 3. If you are using big data, are you using it in your marketing?**

       a) Yes
       b) No, why?

Big data is used in marketing



■ Yes ■ No
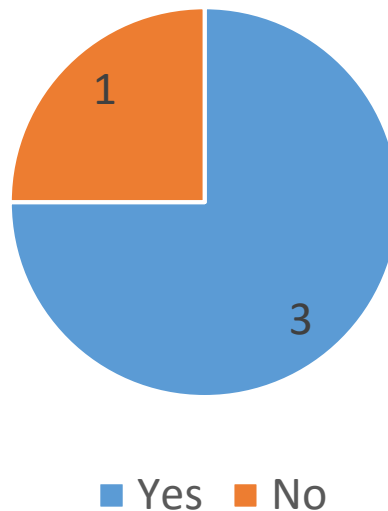
*Figure 3.2 This graph visualises the answers of does the companies use big data in their marketing.*

The distribution for this question were 3 for "yes" and 1 for "no", only 4 answers in this section since one of the companies declared that they are not applying big data. The ones (3) who answered *yes*, are using big data for discovering customer segments, maybe even on store level. Customer behaviour patterns are created, and loyal customer marketing is personalized. The one who answered *no*, may use even 20 different channels of data but are not connected together so could not be characterised as big data. Even if the excluded company from this chart was not applying big data, they used data for example in advanced analytics for supporting marketing campaigns.

**Question 4. Do you base decision making purely on data?**

Decision are made purely based on data



Yes  No
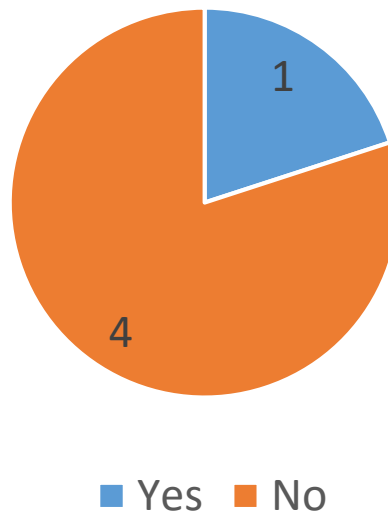
*Figure 3.3 This pie chart shows the amount of companies that makes decisions purely based on data. 4 of the participants said "no" and only on answered "yes".*

The most of the answers stated that decisions cannot be made solely based on data, and needs a human judgement no make the final call. Only one organisation has decision making tools purely based on data. Although, all of the respondents stated that data is strongly present in decision making.

**Question 5. How do you collect data, what kind of sources do you use, external/internal?**

Internal / external sources used for data collection



Internal  External  Both
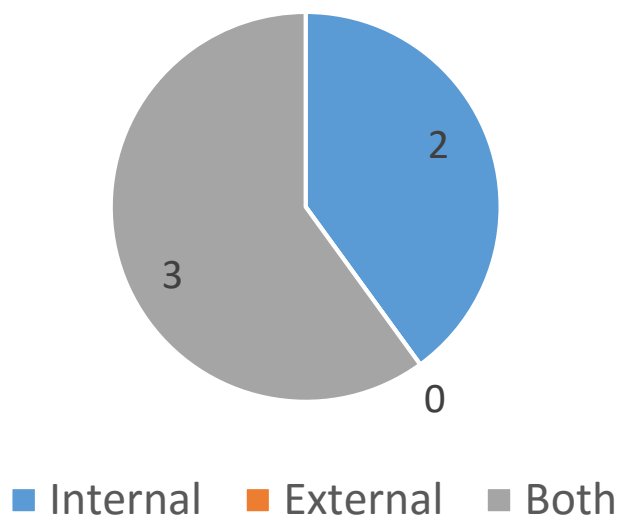
*Figure 3.4 The distribution between different sources used for data collection among the companies.*

3/5 of the respondents answered that both sources are used, and 2/5 participants only used internal sources. Although the distribution was moderate, most of the organisations stated that internal sources are used more broadly, even if they used both sources. External sources were used for example in mapping consumer behaviour in the web. Internal sources collect information of the regular customers, customer behaviour, demographics, campaign effectiveness on average purchase behaviour and also satisfaction levels with customer questionnaires.

**Question 6. Are you using your own maintained physical, or cloud based databases?**

Data storage solution



■ Own physical databases    ■ Cloud solutions    ■ Both
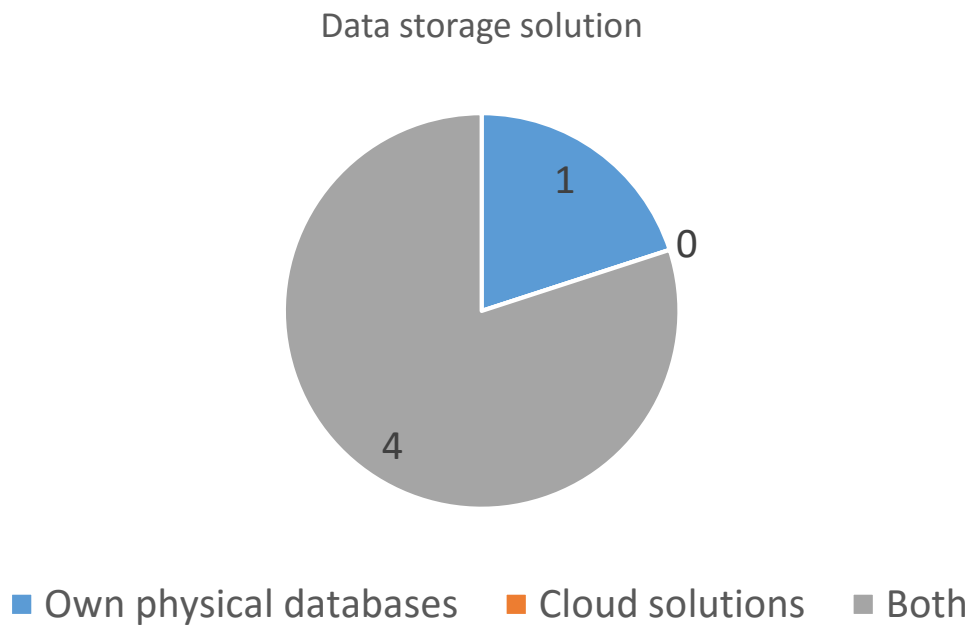
*Figure 3.5 How is the data being stored among the participated organisations.*

The majority (4/5) uses both solution and one participant uses only their own physical database. The graph gives a strong suggestion on the distribution of the data storage methodology. Under the surface the results were quite diverse from each other. Some of the organisations were trying to apply most of the data to the cloud, some wanted to stay working only with own physical databases, and some wanted to apply them both. Different from the graph, it is hard to build any trend among the answers, which of the solutions could be the most popular at the moment or what is the future way. The differences between the data storing could be in the industry and the type of data these varying organisations collect and maintain.

**Question 7. Have you observed whether you collect any dark data (unused data)?**

Is dark data being collected?
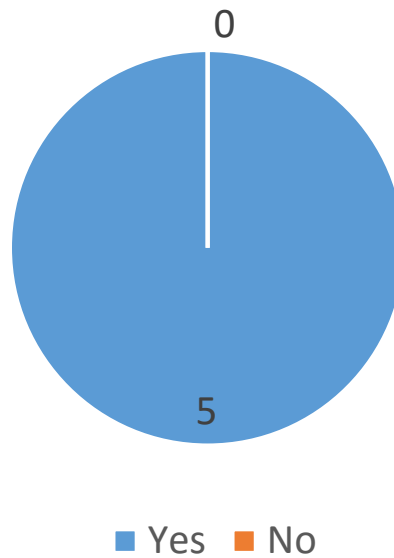


*Figure 3.6 Every one (5/5) of the participants acknowledged that dark data is being collected in some level.*

As visualised, all of the respondents in fact acknowledged that dark data is being collected. There were some differences in the knowledge, how much of this so called dark data is being collected. The amount of dark data varied from unknown, to between 5% and 75% of all the data, which results a huge difference between the companies. There was only one organisation that did not specify the average amount of data being collected, and as said, the other answers varied a lot. Altogether, the majority (3/5), suggested that the amount of dark data takes more than 50% of all the collected data.

**Question 8. How does your company benefit from using big data in their marketing? Name 3 points**
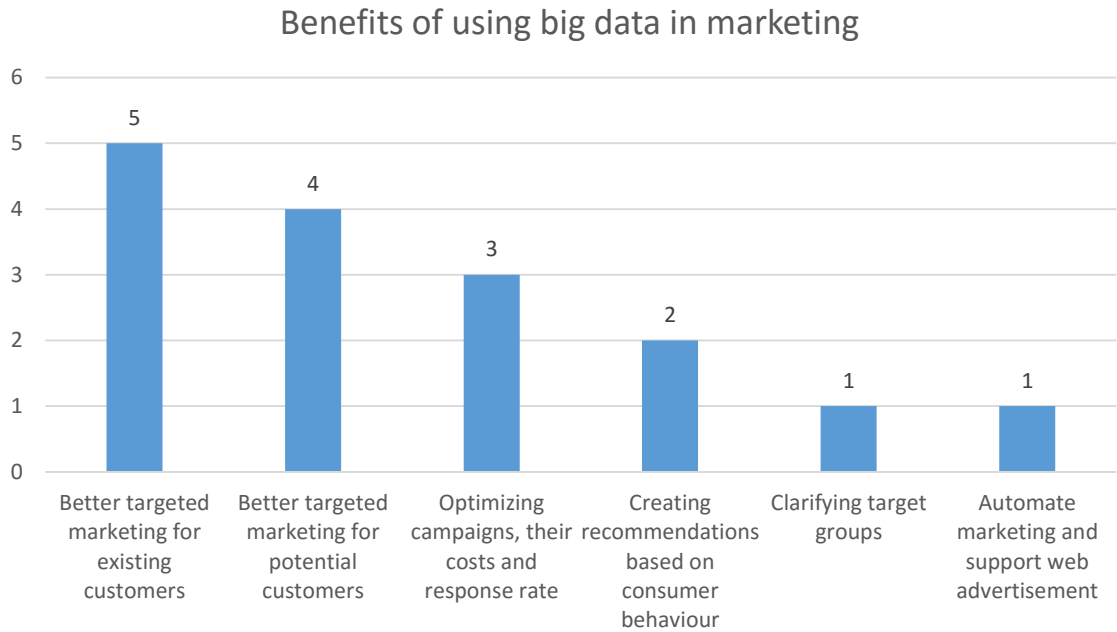


*Figure 3.7 The main findings when asking 3 key points how to benefit from big data usage in marketing. The axis in the left side indicates the number of participants who was to express the need for a benefit described in the bottom axis.*

The question about benefits of big data used in marketing generated a fair amount of similar answers among the organisations. As demonstrated in the graph, most of the organisations gave "better targeted marketing for existing customers" as their first choice of benefits from big data in their marketing. The most common second answer was "better targeted marketing for potential customers". This creates a strong foundation for assuming that at the moment, organisations want to have better target marketing throughout their business field.

**Question 9. How much analytics are included in your marketing from your opinion?**

    a) Very much
    b) Quite a lot
    c) Some
    d) Not that much
    e) Not at all
    f) Something else, what?
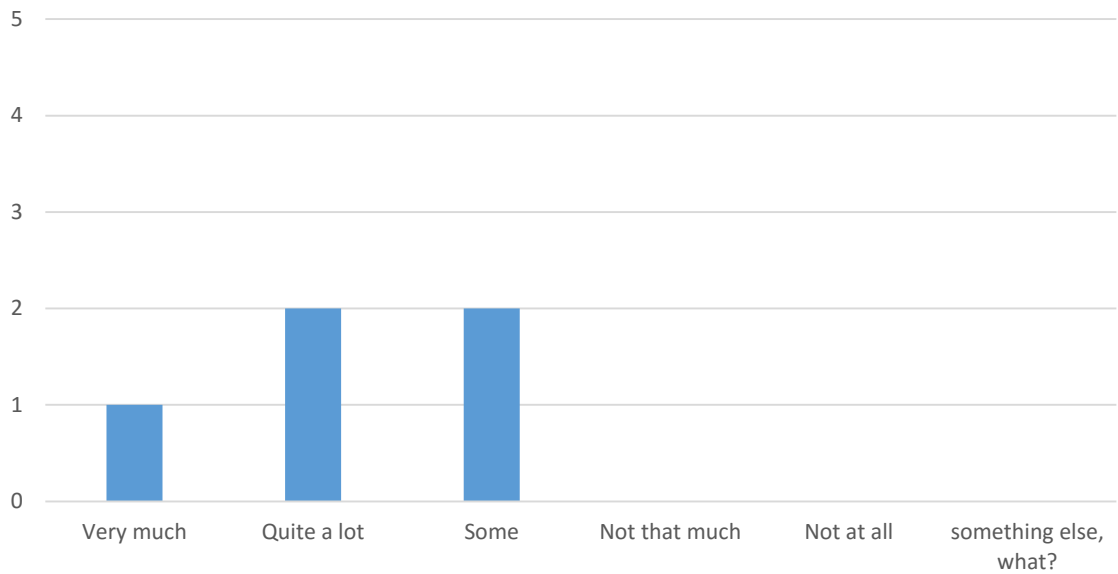
## Amount of analytics included in marketing



*Figure 3.8 The graph shows that how the companies express the level of analytics used in their marketing. The left axis shows the amount of companies and the bottom axis shows the level of usage.*

The answers for this questions shows the present state of the companies´ analytics, most of them stated that it will grow towards the future in fairly rapid fashion. Which indicates that the usage of data is growing, since analytics are based on data.

**Question 10. Do you use predictive analytics in your marketing?**

## Predictive analytics used in marketing
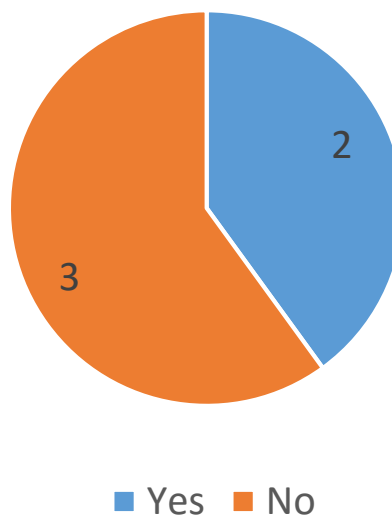


■ Yes ■ No

*Figure 3.9 The chart shows the usage of predictive analytics in the marketing of the interviewed companies. The number and distribution of "yes" and "no" answers are displayed above.*

The results for predictive analytics are fairly ambiguous. The organisations that used predictive analytics states that it is used in several ways, but the ones that said that are not using predictive analytics, might have had some tests on it, or have some kind of predictions still made. Only one of the organisations stated a clear "no" to the question. The graph gives a clear answer on the usage but it is not that simple when going deeper. The majority of the organisations are not using predictive analytics authentically, but most of them are working with it in some level.

**Question 11. What kind of downsides do you see in using big data in your marketing?**

Downsides of using big data in marketing
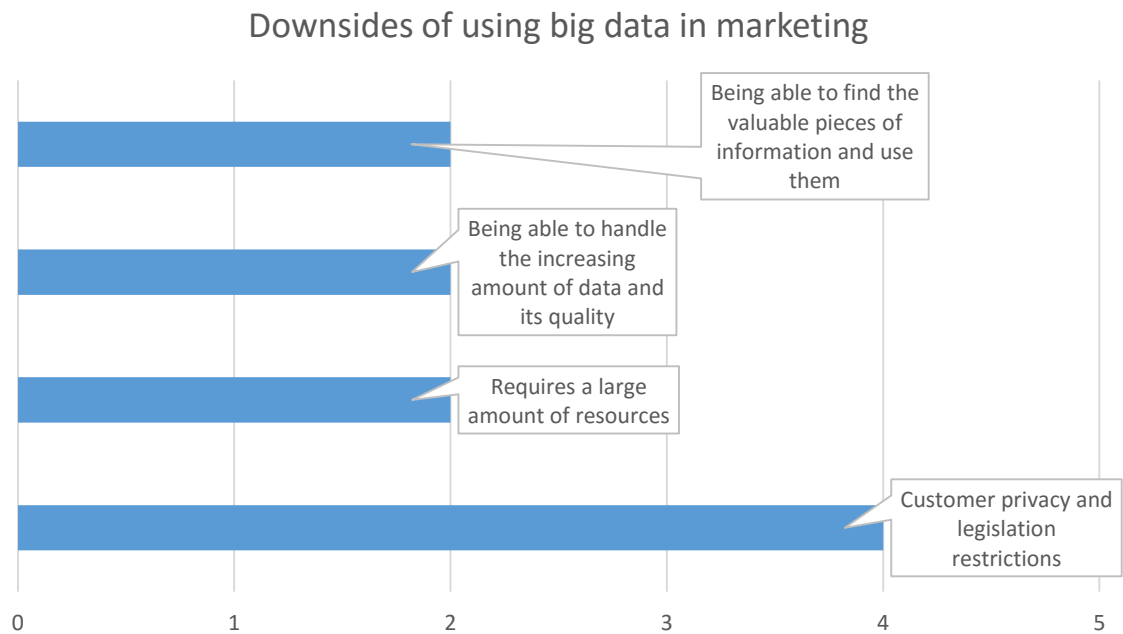


*Figure 3.10 The chart displays the downsides of big data usage in marketing. The bottom axis indicates the number of companies and the data labels expresses the downside.*

As the graph presents, most of the downsides of using big data were related in legislation problems and the customer permission on data collection. It seems that the organisations are struggling with the fact that they cannot collect or use the collected data from the customers without interfering their personal space.

**Question 12. How well would you describe your company knows their customers (segments)?**

| Loyal and regular customers |
|---|
| • Are known very well among all of the participants |

| Occational customers |
|---|
| • Not enough information from occational customers |

| All potential customers |
|---|
| • One organisation has good general knowledge on all the customers |
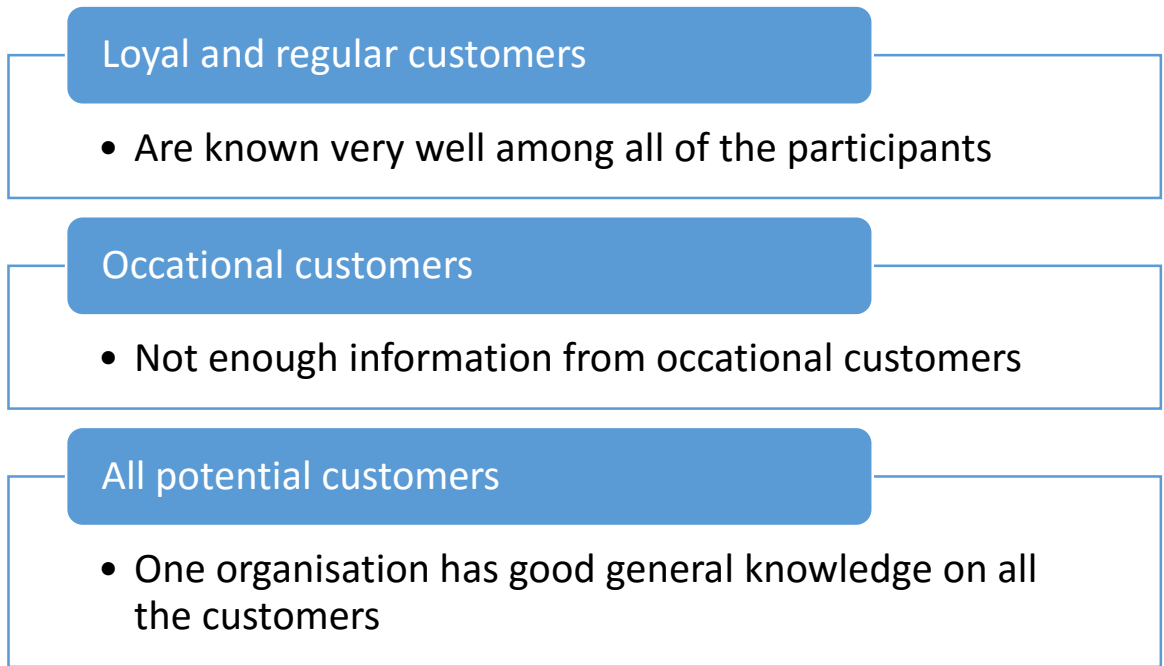
*Figure 3.11 The figure divides the customer base in 3 categories shown above, the most common answers are presented as the level of knowledge on each of these categories.*

All of the companies declared that they know their regular and loyal customers very well. The biggest problem seems to be in understanding and identifying the customers that occasionally purchase products or services from the company. Although, some of the companies stated that they have fairly good understanding of their potential customers as well. In general, the companies collect data from random customers but the problem is that they cannot identify and personalize them, which means the data cannot be connected to different purchasing events and therefore does not bring insights of the customers to the organisation.

**Question 13. What kind of weight / distribution has target marketing in your marketing mix?**

Distribution of target marketing in the marketing mix
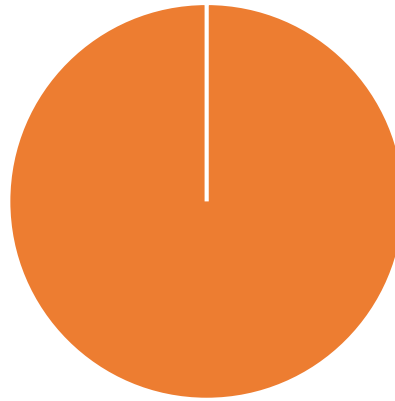


■ More than half of the marketing is targeted

■ Less than half of the marketing is targeted

*Figure 3.12 The chart displays the distribution of target marketing from the whole marketing efforts of an organisation, which results that every one of the organisation uses more resources in untargeted than targeted marketing.*

The result shows that all of the companies use more resources on untargeted marketing than they do in targeted. However, in most of the companies, the distribution is very equal between the two. In 4 out of 5, the distribution is around just over 50%, which means the amount of resources used for untargeted marketing is not that much greater. And in fact, some stated that the trend is moving towards targeted marketing, which might take the larger share in the future. All in all, the amount of resources used for untargeted marketing is higher, but only a bit. There was only one company who estimated the distribution to be 80%-20% in favour of untargeted marketing.

**Question 14. Do you think big data usage will change significantly in the near future (5 years)?**
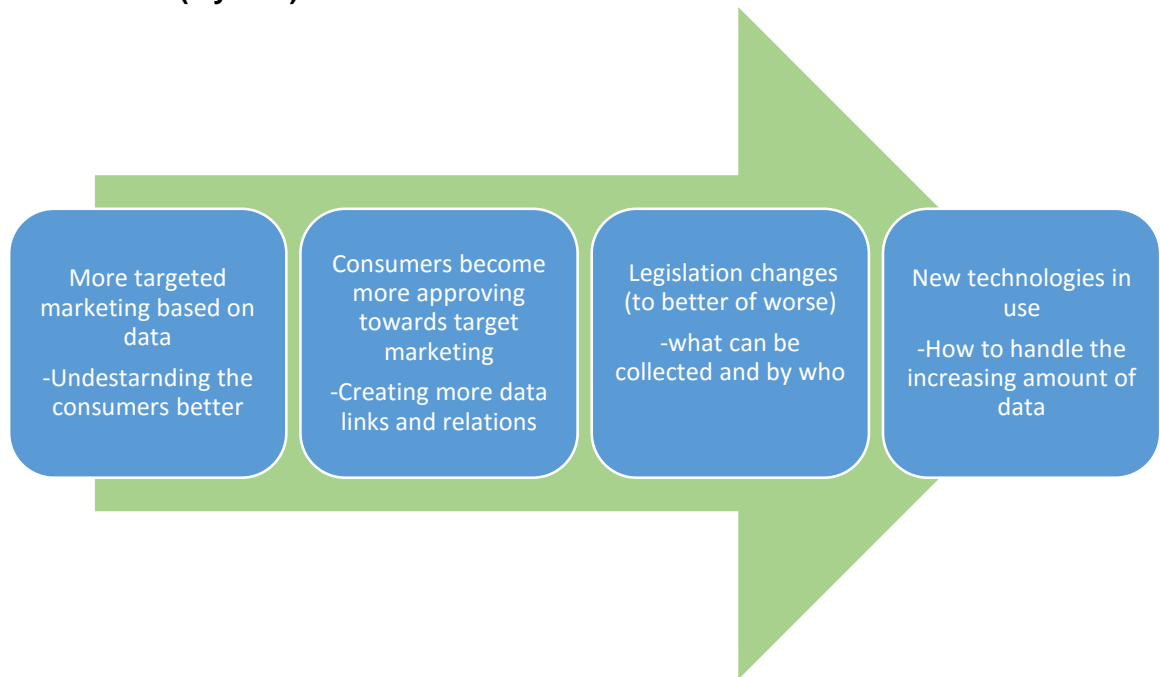


*Figure 3.13 The visualisation presents the interviews´ most occurred thoughts for future of big data usage.*

All of the respondents agreed that the future will bring changes to big data usage. Most of them mentioned the legislation issues that might take part in the data collection and have an effect on being able to create a more personalized view of the customers. Customers might behave more favourable towards the data being collected from their actions. This would enable the companies to create even more personalized marketing, and reach the customers in more various environments. Also big data could able more optimized operations, pricing and campaign management in the future, and allow analytics done in real time. New data relations could be drawn from the collected data, and by this way understand the customers more throughout. Some other negative perspectives were stated added the legislation issues; it could be possible that the amount data grows so big that it makes it harder to concentrate on the right pieces of data. Also it could harden the processing and analysing of the data. The increasing amount of data could result on low quality of data, which might create bugs in the database for example.

# 5   Discussion

The validity of the research was discussed earlier and chosen to be true in some extent. The way, the interviews were conducted and the size of the sampling can be considered to be enough to create a generalization of the results. The main objective of the research was to find out, whether Finnish business-to-consumer companies use big data in their marketing. The answer is yes. Although, any generalisation about the total situation of how many companies does, or does not apply big data technologies in their marketing cannot be made. The majority of the participants of this research were using big data in their marketing, but it does not mean that the same distribution would be true when considering the whole nation.

## 5.1.1   Benefits and downsides

The other objectives of the research were to find out what is big data used for and what kind of benefits and downsides does it bring to the companies. Whether the companies see their present big data usage is on a good level and would it possibly change in the future. The first question for scenario 3 was; what kind of benefits are gained from using big data in marketing. This question was opened in the results and the most common answers were related to target marketing, if divided, targeting marketing for existing customer and for potential ones after that. All of the companies concentrated a lot on their regular and loyal customers and had a good knowledge and understanding of them.

The second most answered benefit was, to be able to create target marketing for the occasional customers. Yet, this was also stated to be a problem. The companies declared the importance of creating targeted marketing for the occasional customers, but they are not able to collect the needed data from these customers, or use the collected data. This suggests that a major milestone for the organisations would be to find a way to collect, or use the information from the occasional customers and create a target marketing solutions based on that data. The most of the organisations suggested that the future might bring some changes to the legislations and perhaps make it easier to collect data from the customers. This being said, it is possible that the direction is totally different, legislation for collecting of data might be even stricter in the future. Another angle to the data collection could be made as well, if the companies are not able to collect data from the occasional customers, perhaps the existing regular customers could act as a base of information. Behavioural patterns could be created from the well-known customers, and apply it to the occasional customers.

As discussed, the most common downside of the usage of big data was related to legislation problems and privacy of consumers. These problems though are less dependent on the organisations itself, and cannot be significantly influenced by them. Other than introducing the data methodologies to their customer, so they could understand what is the data used for. By doing this, maybe the organisation could create a better understanding and trust between the customers. Perhaps this way the customers might become more receiving and approving towards the data used of them, and the companies could truly unchain their analytics used of the customers and create even more targeted marketing.

### 5.1.2   What is big data used for?

As stated, majority of the companies defined that the most benefits come from target marketing, but other operations came up too. Some of the companies told that they use big data for optimizing campaign costs and response rate for example on catalogue distribution. Big data is used in creating recommendations based on customer behaviour, and automating marketing operations. Big data is also being used to track the consumers web behaviour, from which site they clicked the link to visit the particular company´s website for example. This way a link between different sites can be made and also increase the amount of information of the consumer, their preferences and such.

The results from the interview question number 2 showed that none of these particular companies saw their level of big data usage in the marketing being "very good". All of the companies acknowledge some room for improvement on various topics, but are trying to solve them in some level. The general level of big data usage in Finland was discussed in the theory part and the interview results suggest the same situation, Finland is well behind of the leaders. The reason for this could be that big data technologies have been applied for such a short period of time that the companies have not yet reached the desired level. The atmosphere towards big data was quite similar in the companies, it is being applied in small steps and the attitude towards the technology is quite careful. The companies do not want to rush into the usage of big data, which is understandable and reasonable. In the end, all the companies stated that their big data usage is increasing and more target marketing will be done in the future, than today. Also, new technologies will be introduced to create new ways of using big data.

All of the organisations used analytics and slightly over half used predictive analytics in their marketing. Analytics were used in different levels, some used analytics more than others, as it was presented in the question 9 and corresponding figure 3.8. The predictive analytics used by 3 out of 5 indicated that the technology cannot be taken for granted. Not all the companies are including predictive analytics in their organisation. Predictive analytics were

used for example in campaign management and target marketing in the way that the previous campaigns effectiveness was analysed and then improved to the next one if possible. Regarding target marketing, the amount of occurred target marketing was compared to customer satisfaction. Which means that the organisation tries to find out the optimal amount of marketing that can be targeted for an individual, and on what point do they feel it is intervening their privacy or become too overwhelming.

Major differences occurred in the share of dark data, among all the collected data. As stated in question number 7, all of the companies acknowledged there are some unused data within the collected entity. The amount varied a lot between the companies, one of the companies stated that there is only 5% of dark data from all collected. Three of the companies said it is more than 50% all the data, from which one estimated it being around 75%. One company stated that they do have dark data but could not estimate how much does it comprehend from all the data. This concludes the differences of the utilisation level of the collected data, but any generalisation cannot be made who has better utilisation tools for example. The amount of collected data of all the companies are high but still when the amount of data is unknown, the data utilisation cannot be truthfully compared. One statement that could be made from this information, is that some of the companies are able to utilise a larger share of their collected data, than others. Another statement that can be made based on these results is that some of the companies are collecting more of such data that they can utilise at the present state of their operations. Moreover, these companies could be considered to be more efficient on realising their needs of data, and how they can create value from it. Then again, no major connections were found with the amount of collected dark data and the big data usage level in the marketing of the company. This leaves the dark data topic slightly open, and would require further inspection to have more solid base for the assumptions.

### 5.1.3 What does it require from a company?

Big data utilisation itself did not seem to be a major problem in sense of resources, since only one company stated it as a downside. One organisation mentioned that the knowledge is one the main issues, there has to be people who have knowledge and experience of the technology, and big data is not that simple to implement into an organisation. This topic came up in the theory part as well and was characterised as one of the key elements for Finland, and not meeting the needed level of big data know-how. The knowledge is being built along the way with the slow implementation of big data within the organisations, in some more rapidly than in others.

### 5.1.4 Development for further research

If the research was to be continued, the topic scope could be wider. The sampling should be increased, depending on the schedule and the scope. The questions could stay the same if the subject would be kept in the same depth as in this research. The research has potential on larger scale if the amount of participants would be increased. In my opinion with small changes, for example the data storing part, it could be used to generate a good overview of the general level of big data in the marketing of large Finnish B2C companies. The data storing question could be totally removed from the question list, unless the scope for further research would be concentrating in specific field of industry. This way the reliability and the validity of making generalisations would be maximized, when including companies with similar data, and require similar storing methods. Another point for further research would be in the dark data section, the companies were not able to tell how much data they have at the moment, but somehow relate the amount of dark data in some other factor. This way maybe a bit more relevant and reliable comparing could be made between the companies, and who, for example, has utilised the collected the data better than others.

### 5.1.5 Thesis work evaluation

This particular research done with the amount of participants and the level of the questions was valid, as discussed earlier. The depth of the research can be considered suitable for a thesis. The sampling size is debatable, for this research it was valid. The questions used in the interviews worked well, after the results some of them were discovered to be somewhat irrelevant. The different industries and their data differ in such ways that it is hard to compare them with each other, for example the question of data storage, whether the organisation uses internal or external databases. This question was not suitable for creating a good generalisation of the organisations since some of the data was very sensitive, which affects the storing methods. Other than that, the questions worked well for their purpose. It was possible to create a general view of the companies´ big data usage in their marketing in suitable extent. Another point of view for the validity of the research is relate to the dark data analysis part in the discussion, on what big data is used for.

The process for the thesis work was well planned and executed, I started with the theoretical part and created the questions along the way. Before finishing the theoretical part, I started arranging interviews, so the process had a logical flow. I could have started the interview arranging earlier, since I had set up a goal for 10 interviews. I managed to complete 5 interviews and I was told that it can be considered valid, so the arranging of interviews was stopped. Also, the final interview was conducted just 2 weeks before the deadline, and the report needed to be finalized.

The thesis process taught me time management probably more than anything else, if not considering the information and knowledge I gained from the subject. At the beginning of February when I started the project, I started from the sources. I did some research on the literature I could use and got a pile of books related to big data, analytics, marketing, customer behaviour, databases, and other books that might bring in something valuable to the table. After I had the books I decided to write around 2 pages of theory every day I had the chance to. I used this method for 2 different reasons: be able to finish the thesis in time and to keep the subject in my mind all the time, this way I was able to continue the thesis every time from a similar state of writing and create similar text around all the different subjects.

If considering the subject, and what I learned from big data and all the other subjects included in the theoretical framework. I gained a lot of knowledge about the state of big data usage of today, and understood specific types of usages for the technology, and of course the challenges of it. I also got to know marketing in such way that I have not known before, marketing has not been my field of business, or related to my studies. This was a good opportunity for me to broaden my knowledge on a subject that I am interested in.

# References

Adrian, M., Beyer, M., Buytendijk, F., Chandler, N., Heudecker, N., Laney, D., Linden, A. & White, A. 2014. Answering Big Data´s 10 biggest vision and strategic questions. Gartner. (https://www.gartner.com/doc/2822220?refval=&pcp=mpeu) Website accessed 19.02.2016

Alanko, M & Salo, I. 2013. Big data Suomessa. Research for ministry of transport and communications. (http://www.lvm.fi/documents/20181/799435/Julkaisuja+25-2013/5d025b48-038f-49de-bfee-70b2c760c7b7?version=1.0). Website accessed: 3.3.2016

Arthur, L. 2013. Big Data Marketing: Engage your customers more effectively and drive value. Hoboken. New Jersey.

Deshpande, B. 2011. 2 main differences between classification and regression trees. (http://www.simafore.com/blog/bid/62482/2-main-differences-between-classification-and-regression-trees). Website accessed 26.2.2016

Fahy, J & Jobber, D. 2012. Foundations of marketing. 4th ed. Maidenhead. Berkshire.

Finlay, S. 2014. Predictive analytics, data mining and big data: myths, misconceptions and methods. New York. NY.

Franks, B. 2012. Taming the big data tidal wave: finding opportunities in huge data streams with advanced analytics. Hoboken. New Jersey.

Gartner. 16.09.2015. Gartner surveys shows more than 75 percent of companies are investing or planning to invest in big data in the next two years. Gartner newsroom. (http://www.gartner.com/newsroom/id/3130817). Website accessed: 5.3.2016.

Grigsby, M. 2015. Marketing Analytics: a practical guide to real marketing science. London. United Kingdom.

Hovi, A., Hervonen, H. & Koistinen, H. 2009. Tietovarastot ja business intelligence. Jyväskylä. Finland.

Marr, B. 2015. Spark or Hadoop – Which is the best big data framework? Forbes. (http://www.forbes.com/sites/bernardmarr/2015/06/22/spark-or-hadoop-which-is-the-best-big-data-framework/2/#31560d625dfd). Website accessed 26.2.2016

Press, G. 2013. A very short history of big data. Forbes: (http://www.forbes.com/sites/gilpress/2013/05/09/a-very-short-history-of-big-data/#799f3d4355da). Website accessed 9.4.2016.

Press, G. 2015. 6 observations from a new survey of the state of Big Data analytics. (http://www.forbes.com/sites/gilpress/2015/09/04/6-observations-from-a-new-survey-on-the-state-of-big-data-analytics/#2b0b13c4777e). Website accessed 23.2.2016.

Pursiainen, S. 2015. Big data marketing in Finnish startups. Research done for Research Seminar course in Haaga-Helia. Not published.

Reh. J. Pareto´s principle – the 80-20 rule. (http://management.about.com/cs/generalmanagement/a/Pareto081202.htm) Website accessed 8.2.2016.

Salo, I. 2010. Cloud computing: palvelut verkossa. Jyväskylä. Finland.

Salo, I. 2013. Big Data. Tiedon vallankumous. Jyväskylä. Finland.

Salo, I. 2014 Big Data & Pilvipalvelut. Jyväskylä. Finland.

Salo, I. 2015 Big Data, Big Data-asiaa vuodesta 2011 – blog (http://www.bigdata.fi/hadoop). Website accessed 12.2.2016

Salo, I. 2016 Big Data vuonna 2016. (http://www.bigdata.fi/artikkelit/big-data-vuonna-2016) Website accessed 6.3.2016.

Schiffman, L. & Kanuk, L. 2010. Consumer behaviour. 10th ed. Pearson education. New Jersey.

Slocum, M. 2013. Big Data in Europe. The organizers of Big Data Week discuss the Europe´s adoption of big data and data science. (http://radar.oreilly.com/2012/04/big-data-in-europe.html) Website accessed 12.3.2016.

Gartner. IT Glossary: Big data. (http://www.gartner.com/it-glossary/big-data/) Website Accessed 12.04.2016

Marketing terms. 2016. B2C. (http://www.marketingterms.com/dictionary/b2c/) Website accessed 13.04.2016

Tollin, K & Caru, A. 2008. Strategic market creation: a new perspective on marketing and innovation management. Hoboken. USA.

Vaakkuri, M. 2013. Big Data muuttaa maailmaa. Talouselämä. Suomi. (http://www.talouselama.fi/kumppaniblogit/tieto/big-data-muuttaa-maailmaa-3440603) Website accessed 23.2.2016.

Weinman, J. 2015. Digital Disciplines: Attaining market leadership via the cloud, mobile, big data, social, mobile and the internet of things. Hoboken. USA. (http://ezproxy.haaga-helia.fi:2181/ehost/detail/detail?sid=c3611334-0333-4f15-9d2d-7887cada253d%40sessionmgr4001&vid=0&hid=4109&bdata=JnNpdGU9ZWhvc3QtbGl2ZQ%3d%3d#AN=1046505&db=nlebk)

Wu, J. 22.11.2000. The transition of data into wisdom. (http://www.information-management.com/news/columns/-2784-1.html) Website accessed 5.4.2016.

FIGURES:

(Figure 1.1) Gessman. M. 04.08.2015. Generalized linear models in R. Mages´ blog (http://www.magesblog.com/2015/08/generalised-linear-models-in-r.html).          Website accessed 11.2.2016.

(Figure 1.2) Deshpande, B. 2011. 2 main differences between classification and regression trees. Simafore blog. (http://www.simafore.com/blog/bid/62482/2-main-differences-between-classification-and-regression-trees. Website accessed 26.2.2016).

# Appendices

## 5.2 Appendix 1. The interview questions

1. What do you consider being big data?
2. How would you describe the usage level of big data in your company?
    a. Very good
    b. Good
    c. Rather good
    d. Rather basic
    e. Bad
    f. We are not using big data
    g. something else, what?
3. If you are using big data, are you using it in your marketing?
    a. Yes
    b. No, why?
4. Do you base decision making purely on data?
5. How do you collect data, what kind of sources do you use, external/internal?
6. Are you using your own maintained physical, or cloud based databases?
7. Have you observed whether you collect any dark data (unused data)?
8. How does your company benefit from using big data in their marketing? Name 3 points
9. How much analytics are included in your marketing from your opinion?
    a. Very much
    b. Quite a lot
    c. Some
    d. Not that much
    e. Not at all
    f. Something else, what?
10. Do you use predictive analytics in your marketing?
11. What kind of downsides do you see in using big data in your marketing?
12. How well would you describe your company knows their customers (segments)?
13. What kind of weight / distribution has target marketing in your marketing mix?
14. Do you think big data usage will change significantly in the near future (5 years)?